

Crowdsourcing Customer Needs for Product Design using Text Analytics

Stuart Palmer

Abstract—In product design and engineering, identifying customer needs is the foundation for designing and producing a successful product. Traditionally, a range of techniques have been employed to elicit customer needs. A relatively new technique for identifying customer needs is ‘crowdsourcing’. An emerging area of research is the crowdsourcing of customer needs from online product review sites. This paper proposes a simple process for crowdsourcing customer needs for product design using text analytics. The analysis/visualization method is presented in detail. The text content of online customer reviews for a popular product is collected and processed using text analytics software. A published case study identifying expressed customer needs for the same generic product, collected via conventional means, is used to successfully validate the findings from the text analytics method.

Index Terms—crowdsourcing, customer needs, product design, text analytics, customer reviews

I. INTRODUCTION

IN product design and engineering, identifying customer needs is the foundation for designing and producing a successful product [1], [2]. Traditionally, a range of techniques have been employed to elicit customer needs, including surveys [3], [4], interviews [2], [5], focus groups [6], [7], market research [5], [8], interactive queries [4], and examining sales data [6]. However, many of these techniques are labor intensive, time consuming, and expensive [9]. Data obtained via structured customer needs elicitation processes, such as surveys and focus groups, can suffer from systematic validity issues. A closed set of questions may not discover all customer needs [7], and a formal elicitation process may not gather genuine customer responses if a respondent believes that their opinion may be unpopular [10].

A relatively new technique for identifying customer needs is ‘crowdsourcing’. Many contexts for, and definitions of, crowdsourcing exist. It might be, “... outsourcing a task to a large group of people ... (a crowd) ...” [11, p551], or, “... a process that enables and facilitates acquiring knowledge and experience.” [8, p20] In a product design/development context it might refer to, “... products whose definition and/or specification and/or design occur with multiple direct consumer inputs ...” [12, p5]. Examples of the use of crowdsourcing in product design include:

- 1) Employing online task markets such as Amazon’s Mechanical Turk platform to source public

- contributions for design-related tasks [4], [13].
- 2) Design competitions open to the public [14], [15].
- 3) Seeking new product ideas from online customer communities cultivated by corporations [8], [14], particularly seeking the views of so called ‘lead users’ who are typically early adopters or ‘power users’ of products [16], [17].
- 4) Using social media platforms to seek/find feedback from customers on existing or proposed products [10], [11].
- 5) Open design/open innovation platforms that allow a distributed crowd of contributors to collaborate on design projects [18], [19].
- 6) Engaging customers and/or others in product design evaluation refinement activities [4], [20].
- 7) Innovation crowdsourcing forums within an organization [21].

Crowdsourcing product design information via social media would typically involve some intentional communication between those seeking the information and those providing it. However, an emerging area of research is the crowdsourcing of customer needs from online product review sites [2], [5]–[7]. There is a wide and growing range of online sites containing customer reviews of both specific classes of products and specific brands of products. However the largest online review site is probably the amazon.com online marketplace. Hence many researchers have focused their efforts there as a source of data for evaluating the utility of determining desired product characteristics from online customer product reviews [1], [3], [10], [22].

A number of benefits are observed in using online customer reviews as a source of product design data. Large quantities of data can be collected with relative ease, continuing timely data is typically available as new reviews are often frequently added, and review data are based on ‘revealed customer preferences’ derived from post-purchase experiences rather than a hypothetical purchasing scenario from a survey [7]. Compared to structured processes such as surveys and focus groups that may constrain the responses obtained, online customer reviews are typically in a free text form and with no constraints on customer sentiments [9], and they are generally provided with no incentive and without the pressure that may be present in a more formal process [10].

There are potential limitations to using online customer review data for identifying customer needs for product design. The public nature of such online sites means that they can be used to spread false or spam product reviews, and online customer reviews generally relate to the features of existing products, whereas it may be a goal of the

Manuscript received February 4, 2016; revised March 10, 2016.

S. R. Palmer is with the Faculty of Science, Engineering and Built Environment, Deakin University, Geelong, VIC 3217 Australia (phone: +613-5227-2991; fax: +613-5563-3462; e-mail: spalm@deakin.edu.au).

customer needs determination process to include speculation regarding future desired product characteristics [23]. While collection of data from online customer review sites may be relatively easy, the analysis of such free text data can be very complex [1], [2], [10]. For particular types of product, it may be that there are limited or no online customer reviews publicly available.

Text analysis may be performed manually [9], but as soon as the amount of text becomes large, computer-based analysis is generally the most practical option [2], [22]. The computer-based analysis and visualization of textual data goes by various names, including text analysis [6], [24], opinion mining [10], [22], text mining, [2], [7], and text analytics [25]. The latter term is used here as the general name for describing, "... a set of linguistic, statistical, and machine learning techniques that model and structure the information content of textual sources for business intelligence, exploratory data analysis, research, or investigation." [25, p388]

This paper proposes a simple process for crowdsourcing customer needs for product design using text analytics. The analysis/visualization method is presented in detail. The text content of online customer reviews for a popular product is collected and processed using text analytics software. A published case study identifying expressed customer needs for the same generic product, collected via conventional means, is used to validate the findings from the text analytics method. Additional desirable future research is also identified.

II. METHOD

The literature on product design was reviewed to identify documented case studies of customer needs determination for specific products. Concurrently, the amazon.com online marketplace was consulted to determine if products similar to those in the identified case studies from the literature were on offer for sale, and whether a significant number of customer-contributed reviews for those products were available. A detailed documented case study of customer needs determination and plentiful online customer reviews were found for a violin stand product [26]. Based on the identified case study, a non-ranked, benchmark list of unique desired product characteristics was established.

In [10], the authors note that text analysis generally consists of three stages: i) data retrieval and preparation; ii) text processing; and iii) analysis. The text analytics software package KH Coder [27] was used to process and analyze the customer review data. KH Coder was selected as it is free and provides a range of analysis and visualization options described below.

All publically available customer reviews for a number of popular similar violin stand products available via the amazon.com website were retrieved. Preparation of the customer reviews included:

- 1) Removal of any sponsored product reviews – where the customer may have received a purchase discount in return for providing a product review.
- 2) Collecting the reviews into a single text file.
- 3) Editing the reviews so that each review was presented as a single paragraph.
- 4) Correcting any obvious typographical errors [2], [17].

- 5) Converting the entire text of the file to lowercase [23].
- 6) Exporting the context of the file in TXT (plain text) format.

For text processing, KH Coder supports the use of a dictionary of ‘stop words’, that is, common words to be ignored in the analysis of the review text because they add little information and would otherwise mask important words/terms due to their high frequency [22]. A stop word dictionary was developed based on the example English stop word dictionary supplied with KH Coder, after inspection to confirm that no words relevant for this analysis were present. Additionally, free-form text, such as customer reviews, are likely to contain terms in related/derived forms, such as ‘cost’, ‘costing’, ‘costs’, ‘costly’, etc. KH Coder supports stemming to convert and consolidate terms into their root form [2], [22], and this process was used.

The first analysis method used was the development of a co-occurrence network (CON) visualization [24]. Co-occurrence refers to the presence of two (or more) terms in the same text unit of analysis – here we are interested if the same term groups frequently co-occur in customer reviews. KH Coder uses the Jaccard distance [25] as a measure of co-occurrence for term pairs. Based on specifying the minimum frequency of occurrence of a term for inclusion in the CON analysis and visualization, terms appear as nodes in a network plot based on the Fruchterman and Reingold layout algorithm [28]. Frequently co-occurring terms in the visualization are connected by lines/edges. It is possible to configure the plot to indicate the relative frequency of terms by the relative size of their node, and to indicate the relative frequency of co-occurrence of terms by the relative thickness of the edge connecting their nodes.

The second analysis method used was the development of a multi-dimensional scaling (MDS) visualization [29]. Generically, MDS computes a measure of ‘distance’ between all pairs of text terms, then seeks a representation of the terms in the least possible number of dimensions, such that original distance values between all term pairs are shown with the least error possible. The error in the resultant visualization is reduced as more dimensions are used, however using more than two dimensions makes the visualization hard to display and interpret visually. KH Coder supports a number of distance measures and dimensional reduction techniques – here we use the Euclidian distance measure [30] and the Kruskal distance scaling method for dimensional reduction [31]. Based on specifying the minimum frequency of occurrence of a term for inclusion in the MDS analysis and visualization, terms appear as circles/bubbles in the plot, and it is possible to configure the plot to indicate the relative frequency of terms by the relative size of their bubble. Here we use two dimensional MDS as a trade-off between the fidelity of the representation of term distances and the ease of interpretation of the visualization. Words/terms clustered close together in the resultant MDS visualization are found more frequently close together in the source text, and may reveal key themes in the customer reviews. Depending on the nature of the source text it may be possible to attribute an ordinal or other meaning to the dimensions of the resultant MDS visualization [29], but here we use the visualization primarily to identify clusters of associated

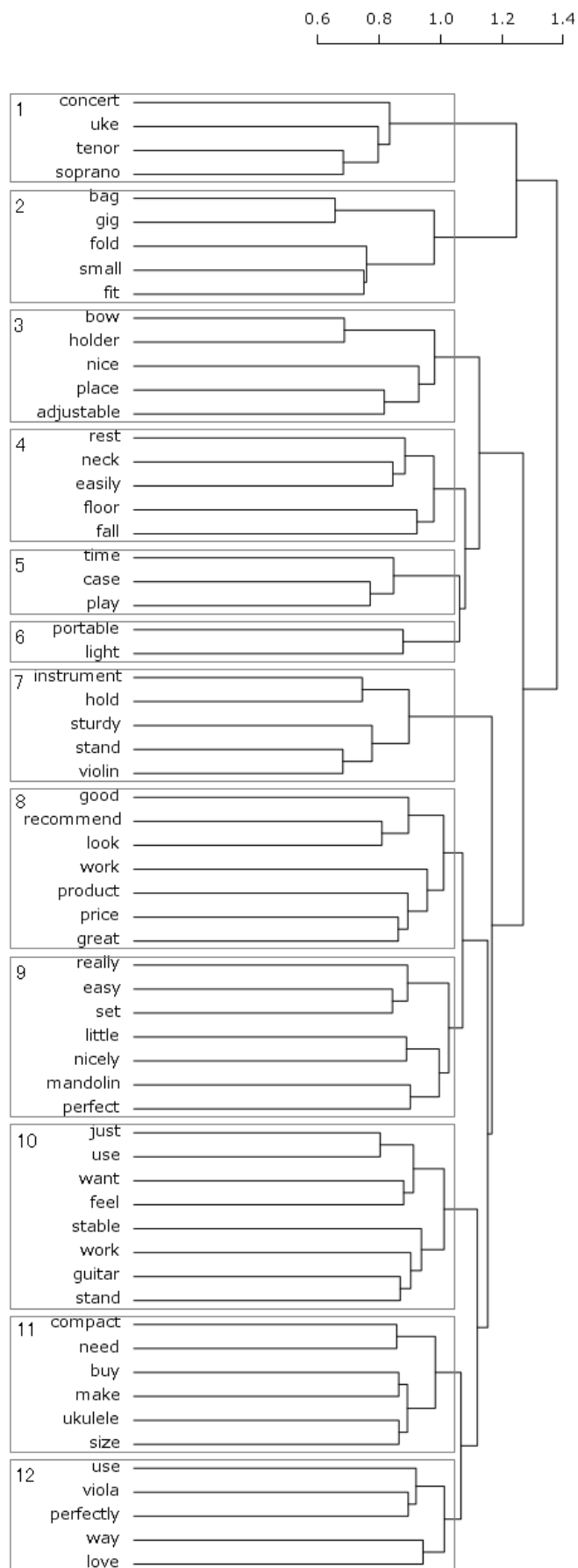


Fig. 3. Hierarchical cluster analysis dendrogram visualization (in 12 clusters) for violin stand online customer reviews, agglomeration dissimilarity coefficient shown on scale at top.

A. General

Each of the visualizations resulting from the CON, MDS, and HCA analyses were examined to find evidence of highlighted themes from the customer reviews corresponding to the six desired product characteristics of a violin stand derived from [26]. Each of the product characteristics is considered in turn below.

B. Stability of Stand/Instrument

In Fig. 1 (CON), the following connected term pairs frequently co-occurring in customer reviews were observed: ‘feel–secure’, ‘heavy–knock’, ‘solid–construction’, ‘sturdy–stand’, and ‘hold–instrument’. In Fig. 2 (MDS plot), the following frequently occurring terms in customer reviews were observed: ‘sturdy’ and ‘stable’, and the closely associated frequently occurring term pair was observed: ‘good–hold’. In Fig. 3 (HCA dendrogram), cluster 7 was observed to include the associated terms: ‘instrument–hold’ and ‘sturdy–stand’, and cluster 10 was observed to include the associated terms: ‘stable–stand’.

C. Low Cost/Good Value

In Fig. 1, the following connected term pair was observed: ‘price–quality’. In Fig. 2, the following closely associated term cluster was observed: ‘great–stand–price’. In Fig. 3, cluster 8 was observed to include the associated terms: ‘great–price–product’.

D. Easy to Use

In Fig. 1, the following connected term pair was observed: ‘easy–use’, and the following connected term cluster was observed: ‘ready–play–time–practice’. In Fig. 2, the following closely associated term pairs were observed: ‘ease–use’ and ‘work–perfect’. In Fig. 3, cluster 9 was observed to include the associated terms: ‘really–easy’.

E. Aesthetically Pleasing

In Fig. 3, cluster 8 was observed to include the associated terms: ‘good–look–recommend’.

F. Can Hold Other Types of Instruments

In Fig. 1, the following connected term pair was observed: ‘violin–stand’, and the following connected term cluster was observed: ‘X–uke–stand’, where X included the following ukulele types: ‘tenor’, ‘soprano’, and ‘concert’. In Fig. 2, the following terms were observed: ‘violin’, ‘mandolin’, ‘ukulele’, and ‘uke’. In Fig. 3, cluster 1 was observed to include the associated terms: ‘uke–concert–tenor–soprano’, cluster 7 was observed to include the associated terms: ‘violin–stand’, cluster 9 was observed to include the associated terms: ‘mandolin–perfect’, cluster 10 was observed to include the associated terms: ‘guitar–stand’, cluster 11 was observed to include the associated terms: ‘ukulele–size’, and cluster 12 was observed to include the associated terms: ‘viola–perfectly’.

G. Portable for Travel

In Fig. 1, the following connected term pair was observed: ‘light–weight’, and the following connected term cluster was observed: ‘fold–small–fit–gig–bag–pocket’. In Fig. 2, the following closely associated term pair was

observed: 'compact-fit', and the following closely associated term clusters were observed: 'little-really-portable' and 'fold-small-bag'. In Fig. 3, cluster 2 was observed to include the associated terms: 'fold-small-fit-gig-bag', and cluster 6 was observed to include the associated terms: 'portable-light'.

H. Other Product Characteristics

The visualizations were also examined for the presence of any other notable themes suggesting additional desired product characteristics implicit in the customer reviews, but not previously identified. In Fig. 1, the following connected term cluster was observed: 'adjustable-place-bow-holder'. In Fig. 3, cluster 3 was observed to include the associated terms: 'adjustable-place-bow-holder'. The text concordance feature of KH Coder was interrogated to confirm that many of the customer reviews made reference to the desirability of a bow holder for a violin stand. While not all small stringed instruments require a bow, the findings here suggested that the provision of an adjustable bow holder is something that should be considered for such a product. Note that this additional desired product characteristic was not identified in the survey-based customer needs elicitation case study presented in [26].

CONCLUSION

This paper proposed a simple process for crowdsourcing customer needs for product design using text analytics. The text content of online customer reviews for a popular product was collected and processed using text analytics software. A published case study identifying expressed customer needs for the same generic product, collected via conventional elicitation means, was used to validate the findings from the text analytics method. All of the six desired product characteristics derived from [26] can be inferred from examining the themes present in the visualizations produced. Additionally, the crowdsourcing/text analytics process identified another desired product characteristic that was not found in the survey-based customer needs elicitation case study used as a benchmark. The HCA dendrogram in Fig. 3 produced the most comprehensive visualization, containing all six of the desired product characteristics, plus the additional desired characteristic. However all three forms of visualization (CON, MDS, and HCA) added value to the analysis presented, providing confirmatory overlapping results for desired product characteristics.

The work presented here was relatively limited in scope, and the following additional research could add value to the process described. Additional analysis methods falling under the broad banner of text analytics, and supported by KH Coder, and which may add value in identifying desired product characteristics from online customer reviews include word frequency analysis [1], and machine learning techniques, such as artificial neural networks [2], [22]. These analysis methods could be applied to the customer review data to produce additional tabulations/visualizations. No attempt was made here to rank the importance of customer needs [2], [3]. Ranking of customer needs may be desirable if the resulting set of product characteristics entails trade-offs. Methods for inferring the ranking of customer

needs from online customer reviews could be investigated. No attempt was made here to identify customer/market segmentation for customer sub-groups with distinct sets of needs [26], [34]. If the market for the product is large enough, it may be worthwhile to attempt to identify customer sub-groups, and provide a differentiated product offering that separately addresses each market segment. Methods for identifying customer sub-groups from online customer reviews could be investigated.

Online customer review sites are growing in number, range of industries/products covered, and quantity of review data contained therein. The process documented here for crowdsourcing customer needs for product design using text analytics is offered as a simple, low-cost way to supplement and validate customer needs data for a product collected by conventional means, where online customer reviews for that product exist.

REFERENCES

- [1] Y. Liu, J. Jin, P. Ji, J. A. Harding, and R. Y. K. Fung, "Identifying helpful online reviews: a product designer's perspective," *Computer-Aided Design*, vol. 45, no. 2, pp. 180-194, 2013.
- [2] R. Rai, "Identifying key product attributes and their importance levels from online customer reviews," in *Proc. ASME International Design Engineering Technical Conference and Computers and Information in Engineering Conference*, Chicago, 2012, pp. DETC2012-70493(1-8).
- [3] J. Jin, P. Ji, and Y. Liu, "Prioritising engineering characteristics based on customer online reviews for quality function deployment," *Journal of Engineering Design*, vol. 25, no. 7-9, pp. 303-324, 2014.
- [4] Y. Ren, A. Burnap, and P. Papalambros, "Quantification of perceptual design attributes using a crowd," in *Proc. 19th International Conference on Engineering Design*, Seoul, 2013, pp. 1-9.
- [5] A. F. Zakaria and S. C. J. Lim, "An instructional approach for product features and customer preference identification using product reviews," in *Proc. National Forum and Exhibition on Engineering Design*, Parit Raja, Johor, 2013, pp. 1-6.
- [6] B. Cornell, I. Kim, and G. E. O. Kremer, "User sentiments in online forums to evaluate and visualize features in multiple-generation products," in *Proc. Industrial and Systems Engineering Research Conference*, Nashville, 2015, pp. 1-10.
- [7] C. Tucker and H. Kim, "Predicting emerging product design trend by mining publicly available customer review data," in *Proc. 18th International Conference on Engineering Design*, Copenhagen, 2011, pp. 1-10.
- [8] A. Chwiałkowska, "Crowdsourcing as a customer relationship building tool," *Journal of Positive Management*, vol. 3, no. 1, pp. 18-32, 2012.
- [9] J. Jin, P. Ji, Y. Liu, and S. C. Johnson Lim, "Translating online customer opinions into engineering characteristics in QFD: a probabilistic language analysis approach," *Engineering Applications of Artificial Intelligence*, vol. 41, pp. 115-127, 2015.
- [10] D. Raghupathi, B. Yannou, R. Farel, and E. Poirson, "Customer sentiment appraisal from user-generated product reviews: a domain independent heuristic algorithm," *International Journal on Interactive Design and Manufacturing*, vol. 9, no. 3, pp. 201-211, 2015.
- [11] M. Bertoni and K. Chirumalla, "Leveraging web 2.0 in new product development: lessons learned from a cross-company study," *Journal of Universal Computer Science*, vol. 17, no. 4, pp. 548-564, 2011.
- [12] M. Sinclair and I. Campbell, "A classification of consumer involvement in new product development," in *Proc. Design Research Society Biennial International Conference*, Umeå, Sweden, 2014, pp. 1-18.
- [13] C. L. Carter and J. A. Edelman, "Emergent design methodologies and implications," in *Proc. 17th International Conference on Engineering Design*, Palo Alto, 2009, pp. 1/551-1/558.
- [14] B. L. Bayus, "Crowdsourcing new product ideas over time: an analysis of the Dell IdeaStorm community," *Management Science*, vol. 59, no. 1, pp. 226-244, 2013.
- [15] D. Wu, J. L. Thames, D. W. Rosen, and D. Schaefer, "Enhancing the product realization process with cloud-based design and manufacturing systems," *Journal of Computing and Information Science in Engineering*, vol. 13, no. 4, pp. 041004:1-14, 2013.
- [16] J. Bragge, T. Tuunanen, and P. Marttiin, "Inviting lead-users from virtual communities to co-create innovative IS services in a structured

- groupware environment," *Service Science*, vol. 1, no. 4, pp. 241-255, 2009.
- [17] S. Tuarob and C. S. Tucker, "Automated discovery of lead users and latent product features by mining large scale social media networks," *Journal of Mechanical Design*, vol. 137, no. 7, pp. 071402:1-11, 2015.
- [18] M. Fuge, K. Tee, A. Agogino, and N. Maton, "Analysis of collaborative design networks: a case study of OpenIDEO," *Journal of Computing and Information Science in Engineering*, vol. 14, no. 2, pp. 021009:1-8, 2014.
- [19] H. Faste, "Opening "open" innovation," in *Proc. 2011 Conference on Designing Pleasurable Products and Interfaces*, Milan, 2011, pp. 1-8.
- [20] L. L. Yu and J. V. Nickerson, "Generating creative ideas through crowds: an experimental study of combination," in *Proc. 32nd International Conference on Information Systems*, Shanghai, 2011, pp. 1-16.
- [21] A. Scupola and H. W. Nicolajsen, "The impact of enterprise crowdsourcing on company innovation culture: the case of an engineering consultancy," in *Nordic Contributions in IS Research*, T. H. Comisso, J. Nørbjerg, and J. Pries-Heje, Eds. Heidelberg New York Dordrecht London: Springer, 2014, pp. 105-120.
- [22] M. Wu, L. Wang, M. Li, and H. Long, "An approach of product usability evaluation based on Web mining in feature fatigue analysis," *Computers & Industrial Engineering*, vol. 75, pp. 230-238, 2014.
- [23] S. Tuarob and C. S. Tucker, "Quantifying product favorability and extracting notable product features using large scale social media data," *Journal of Computing and Information Science in Engineering*, vol. 15, no. 3, pp. 031003:1-12, 2015.
- [24] S. Tuarob and C. S. Tucker, "A product feature inference model for mining implicit customer preferences within large scale social media networks," in *Proc. ASME 2015 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Boston, 2015, pp. V01BT02A002:1-11.
- [25] X. Hu and H. Liu, "Text analytics in social media," in *Mining Text Data*, C. C. Aggarwal and C. Zhai, Eds. New York Dordrecht Heidelberg London: Springer US, 2012, pp. 385-414.
- [26] C. Sotos, G. O. Kremer, and G. Akman, "Customer needs based product family sizing design: the Viper case study," in *Advances in Product Family and Product Platform Design*, T. W. Simpson, J. Jiao, Z. Siddique, and K. Hölttä-Otto, Eds. New York: Springer, 2014, pp. 683-706.
- [27] K. Higuchi, *KH Coder*, 2.00b: Koichi Higuchi, 2015, <http://khc.sourceforge.net/en/>.
- [28] T. M. J. Fruchterman and E. M. Reingold, "Graph drawing by force-directed placement," *Software: Practice and Experience*, vol. 21, no. 11, pp. 1129-1164, 1991.
- [29] E. Namey, G. Guest, L. Thairu, and L. Johnson, "Data reduction techniques for large qualitative data sets," in *Handbook for team-based qualitative research*, G. Guest and K. M. MacQueen, Eds. Plymouth, UK: Altamira Press, 2007, pp. 137-162.
- [30] J. L. Solka, "Text data mining: theory and methods," *Statistics Surveys*, vol. 2, pp. 94-112, 2008.
- [31] L. Chen and A. Buja, "Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis," *Journal of the American Statistical Association*, vol. 104, no. 485, pp. 209-219, 2009.
- [32] C. C. Bridges Jr, "Hierarchical cluster analysis," *Psychological reports*, vol. 18, no. 3, pp. 851-854, 1966.
- [33] R. K. Blashfield, "Mixture model tests of cluster analysis: accuracy of four agglomerative hierarchical methods," *Psychological Bulletin*, vol. 83, no. 3, pp. 377-388, 1976.
- [34] M. Goswami and M. K. Tiwari, "Product feature and functionality driven integrated framework for product commercialization in presence of qualitative consumer reviews," *International Journal of Production Research*, vol. 53, no. 16, pp. 4769-4788, 2014.